DOCUMENT RESUME

ED 218 337 TM 820 378

AUTHOR	Burry, James; And Others
TITLE	Teaching and Testing: Allies or Adversaries.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
REPORT NO	CSE-R-165
PUB DATE	81
NOTE	48p.; Papers presented at a symposium at the Annual Meeting of the National Council on Measurement in Education (Los Angeles, CA, April, 1981).
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Achievement Tests; Administrative Policy; *Educational Testing; Elementary Secondary Education; *Evaluation Methods; *Instructional Development; National Surveys; *School Role; *Test Use; Use Studies
IDENTIFIERS	*Test Use Project

ABSTRACT

This document contains three papers from the Test Use Project of the Center for the Study of Evaluation. A preliminary assessment model is described in the first paper, "Contextual Examination of Test Use: The Test, the Setting, the Cost" by J. Herman and J. Yeh. Empirical findings are sought about the nature of testing and its actual use or non-use in schools. To frame an understanding of testing practices, the model examines various achievement test types used in instructional decision-making. Test characteristics, settings or contexts used, and financial, student opportunity and psychological costs are considered. phase I of the project will culminate in a national survey of teachers and administrators to learn how educators think about and use achievement testing results. Preliminary interviews with elementary and secondary teachers, principals and other school personnel are discussed in the second paper, "The Conduct of Testing from the Classroom Perspective" by D. Dorr-Bremme, C. Lazar-Morrison, and J. Lehman. Findings concern the range of assessment devices and amount of time used in evaluating students. The national survey pilot-test is discussed. In the third paper, "The Design of Testing Programs with Multiple and Complementary Uses" by J. Burry, the three-district preliminary interviews are discussed in the third paper as examinations of educators' views about multiple and complementary uses of assessment in external accountability and instructional decision-making. Design factors such as organizational influences, leadership and policy are considered. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED2183337

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TEACHING AND TESTING:
ALLIES OR ADVERSARIES

James Burry
Don Dorr-Bremme
Joan L. Herman
Charlotte M. Lazar-Morrison
James D. Lehman
Jennie P. Yeh

CSE Report No. 165
1981

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G. Gray
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 378

The papers in this volume were originally presented in a symposium at the Annual Meeting of the National Council on Measurement in Education, Los Angeles, California, 1981.

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California, Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

TABLE OF CONTENTS

	PAGE
Contextual Examination of Test Use: The test, The Setting, The Cost. by Joan L. Herman and Jennie P. Yeh.....	1
The Conduct of Testing from the Classroom Perspective. by Don Dorr-Bremme, Charlote M. Lazar-Morrison, and James D. Lehman	9
The Design of Testing Programs with Multiple and Complementary Uses by James Burry.....	26

CONTEXTUAL EXAMINATION OF TEST USE: THE TEST, THE SETTING, THE COST

Joan L. Herman & Jennie Yeh

There is little doubt that testing in American schooling is increasing in both scope and visibility. Federal program requirements, school board accountability concerns, national and regional assessment needs, state-mandated minimum competency requirements, and the expansion of curriculum-embedded testing programs have increased the amount of testing. A few figures attest to this growth. Kirkland (1971) reported that 75 million standardized tests were taken in 1954 by individuals in educational institutions; Goslin (1963) reported that in 1961 the figure had increased to 100 million ability tests per year. Passage of the Elementary and Secondary Education Act of 1966, with its attendant special programs, clearly led to more standardized testing. Although the exact magnitude is unknown, we do know that a child takes an average of six full standardized achievement test batteries before he or she graduates from high school (Houts, 1975). We also know (GAO, 1975) that at least 90% of the local education agencies throughout the country administer standardized, norm-referenced tests to children within their purview. In addition, 42 states conduct a state assessment program (Kauffman, 1979), and 37 states have adopted minimum competency legislation (Gorth, 1979); such efforts lead to additional yearly testing for students at various grade levels.

As with most highly visible activities, testing also has become the subject of much controversy, and the legal and political systems have entered the debate. Proponents, for their part, have argued that tests

serve a variety of important purposes, can contribute to educational quality control, are an important tool for providing individualized instruction for students, and can contribute to improved educational decision-making. Critics, on the other hand, have decried the arbitrariness of current testing practices (Baker, 1980; Herman & Yeh, 1980), have accused them of bias, and have questioned their appropriateness to the changing functions of education (Tyler, 1977). The quality of available tests continues to be controversial (Hoepfner, et al., 1976; Walker, et al., 1979; Huron Institute, 1978), and moratoriums have been called for (NEA, n.d.).

Despite the great controversy that surrounds testing and its potential uses and abuses, there is little empirical information available about the nature of testing as it actually occurs and is used (or not used) in schools. The Test Use Project at the Center for the Study of Evaluation seeks to fill this gap and answer basic questions about tests and schooling. Phase I of the project is culminating in a national survey of teachers and school administrators.

Clearly, the policy toward testing in this country has been one of accretion, but the full magnitude is undocumented. The CSE Test Use Project was designed to provide such documentation: How much testing is going on in schools? What types of tests are being administered and with what frequency? These are central questions that the study addresses.

To provide a rich description of the testing phenomenon in American schools, the Test Use Project also considers these additional questions:

1. To what extent are tests actually used in schools? Studies a decade ago reported little interest in or utilization of test results (Goslin, 1967). Several more recent local studies similarly report that teachers rely little on the results of standardized tests (Boyd, et al., 1975; Yeh, 1978). What is the current picture of use on the national level? Have newer forms of testing (e.g., minimum-competency, criterion-referenced) influenced patterns of use?

2. What contextual factors influence the administration of tests and the use of tests for instructional decision-making? Previous studies suggest that demographic factors, teacher training, and instructional alternatives affect use. (See for example, Goslin, 1967; Yeh, 1978; Cramer & Slakter, 1968.) Recent research perspectives in measurement, change, and psychology suggest other potentially potent factors.

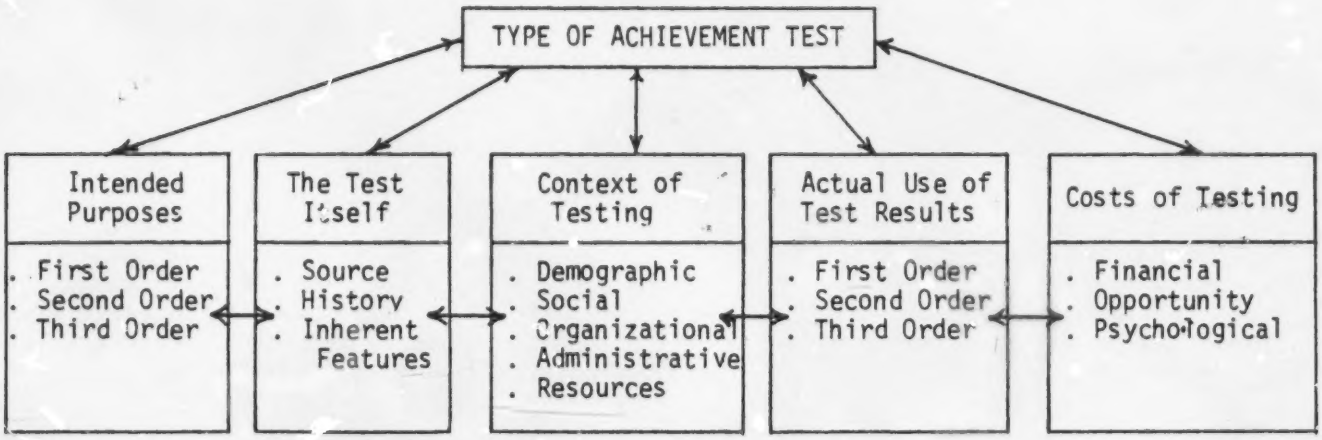
Finally, we felt a coordinate question also must be asked: What does the testing enterprise cost? How much money is spent annually in buying, scoring, and administering formal tests? What other costs, including staff and facilities, are necessary to support testing? Furthermore, where do funds go? What proportion is spent on test purchase, consultant use, computer use, etc.? On the more inferential level, what are regarded as opportunity costs of testing by teachers? What is foregone, and what psychological costs, if any, are imposed? Only by coordinating information about test distribution, the results, and the costs associated with the entire effort can a sounder basis for public policy be developed. Clearly, a sound policy would seek to optimize the utility and minimize the costs of testing.

To bring into better focus the elaborate picture we wanted to frame, a preliminary model was posed (see figure 1). The model suggested that in order to understand testing practices, we need to have, for each type of test administered, some information about the intended purposes, the

characteristics of the test itself, the context of administration, the actual use of results, and the costs. Such a framework enables us to not only describe the nature of testing, but in addition, to explore the relationships between and within the components specified.

FIGURE 1

Framework for Inquiry



The types of tests included within our domain of inquiry were those of achievement, including, for example, standardized norm-referenced tests, criterion-referenced tests, curriculum embedded tests, teacher-made tests, and informal teacher assessments. For the intended, i.e., by the initiator of the test, and actual use of test results, we decided to focus primarily on those uses related to instructional decision-making, e.g., student placement, curriculum planning and revision.

Descriptive characteristics of the test itself included the source of the test, its history, and inherent features. By source, we referred to the process of development and the recency of the test. For example, was

the test developed with broad participation from teachers, community members, and administrators? Was the test developed to measure particular program or curricular objectives? How long has the test been administered? Inherent features of the test characterize the test instrument, for example, test length, ease of administration, specificity of description, perceived validity and reliability, etc. The "Test Itself" component was intended to address the issue of "What is the nature of tests that are currently being administered?"

The "Context of Testing" component addressed the question "In what settings are tests administered?", and included demographic, social, organizational, and resource factors. Demographic factors included such variables as the socioeconomic status of students and the range of special programs at the school site. The social context of testing considered the attitudes of participants, e.g., teachers and principals, toward testing, its utility and importance, and the political environment, e.g., the visibility of test results, and the likely political consequences of those results. Organizational factors included structures for decision-making, and school, district, and classroom organizational patterns that might provide links between testing and instruction, e.g., staff development, grouping patterns. The specific context of administration described factors such as the frequency of testing, and the immediacy of feedback of results. Finally, resources included the district, school, and classroom supports that offer instructional alternatives, e.g., aides, specialists, variety of materials.

The "Cost of Testing" component considered, as already mentioned,

costs of tests, including purchase, development, staff costs, scoring, reporting, etc., at the district and school levels. Opportunity costs were conceptualized in terms of student and staff time, and in activities at all levels that are foregone because of testing. In psychological costs, we were interested in affective consequences for teachers and students, e.g., efficacy, motivation, anxiety, sense of fairness.

This preliminary framework operationalized our initial view of the nature of test practices, and might be used to generate many research hypotheses. For example, given the testing requirements of specially funded programs, it is likely that frequency of testing would be negatively related to socioeconomic status (another context factor). In addition, on the basis of the literature (Goslin, 1967; Yeh, 1978), one might hypothesize that the closer the source of a test to the teacher (a descriptive characteristic), the more likely a teacher would be to use the results of tests for instructional planning.

Obviously, there are a multitude of hypotheses that could be derived from the model, many more than the study could explore adequately. The design phase of the study was intended to narrow the focus, identify the most promising hypotheses, and operationalize better the variables under study. The other papers in this volume discuss some of the results of our work.

REFERENCES

Baker, E.L. Is something better than nothing? Metaphysical test design. In E.L. Baker and E.S. Quellmalz (Eds.), Design, analysis, and policy in testing and evaluation. Beverly Hills, CA: Sage Publications, 1980.

Boyd, J., McKenna, B.H., Stake, R.E., & Yachinsky, J. A study of testing practices in the Royal Oak (Michigan) public schools. Royal Oak, Michigan: Royal Oak City School District, 1975. (ERIC Document Reproduction Service No. ED 117 161)

Cramer, S. & Slatker, M. A scale to assess attitudes toward aptitude testing. Measurement and Evaluation in Guidance, 1968, 1(2).

Gorth, W.P., & Perkins, M.R. A study of minimum competency testing programs: A final report. Washington, D.C.: Office of Testing, Assessment, and Evaluation, National Institute of Education, 1979.

Goslin, D.A. Teachers and testing. New York: The Russell Sage Foundation, 1967.

Government Accounting Office. Problems and needed improvement in evaluating Office of Education programs. Report to the Congress by the Controller General of the United States. Washington, D.C., 1975.

Herman, J.L., & Yeh, J.P. Test use: A review of the issues. In E.L. Baker and E.S. Quellmalz (Eds.), Design, analysis, and policy in testing and evaluation. Beverly Hills, CA: Sage Publications, 1980.

Hoepfner, R., et al. CSE Elementary school test evaluations. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles, 1976.

Houts, P.L. Standardized testing in America II. The National Elementary Principal, 1975, 54(6), 2-3.

Huron Institute on Testing. NCT Staff Circular No. 1, Testing the tests, 1978.

Kauffman, J.D. State assessment program: Current status and a look ahead. Paper presented at the annual meeting of the National Assessment of Educational Progress, Denver, Colorado, 1979.

Kirkland, M. The effects of tests on students and schools. Review of Education Research, 1971, 41(4), 303-350.

National Education Association. Resolution of the National Representative Aseembly of the NEA. Item 1972-28. Washington, D.C.

Tyler, R. What's wrong with standardized testing. Today's Education, 1977.

Walker, C.B., et al. CSE Criterion-referenced handbook. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles, 1979.

Yeh, J.P. Test use in schools. Los Angeles, California: Center for the Study of Evaluation, 1978.

THE CONDUCT OF TESTING FROM THE CLASSROOM PERSPECTIVE

Don Dorr-Bremme
Charlotte M. Lazar-Morrison
James D. Lehman

As part of the work described in the preceding paper, the Test Use Project interviewed forty-four elementary and secondary classroom teachers as well as seven principals and a number of other school personnel to determine how practitioners think about and use the results of student achievement testing. Those interviews were conducted in nine schools across three districts. The interviews attempted to investigate a variety of questions regarding practitioners' use of evaluation techniques in order to aid in the development of the Test Use Project survey instrument that would later be administered nationwide to teachers and proncipals. One of the primary purposes of this preliminary fieldwork was to get an idea of the range of assessment devices being given by elementary and secondary teachers. Another area of investigation was the time teachers actually spend evaluating their students. Some of the results and conclusions that were drawn from the interviews concerning the above questions are presented here.

General Findings

Across the nine schools in the three districts visited, a wide range of assessment techniques was evident. It is important to note, at the outset, that respondents referenced these almost always by their proper names or by vernacular variants of proper names. That is, they rarely talked about "norm-referenced tests," "criterion-referenced tests," "objectives based tests," "curriculum-embedded tests," etc. Instead, they spoke about

"the Ginn placement," "the CTBS," "the Key Math," "that state matrix test," the "Sucher-Allred" and so on. When respondents did refer to kinds of tests, most often they gave them functional class names, e.g., "diagnostic test," "placement tests," "pre-tests," "unit tests," "semester finals," "the competency tests." Exceptions were "standardized tests," "minimum competency tests," and "District tests" (or, the "district testing program," which referred to district-developed, continuum-of-objectives-based measures in the particular sites visited).

These observations are important in that they had obvious implications for our survey instrument development. But they are also noted here to call attention to the fact that the typology of tests and other techniques used in this report is one developed by the researchers using categories salient to the practitioners interviewed.

As expected, a wide range of assessment techniques was reported by the teachers from the nine schools. These 44 teachers (22 elementary and 22 secondary) collectively mentioned the use of eight categories of assessment devices for a total of 351 citations, which is more than likely a low approximation of the actual amount. The assessment categories as well as the number of citations of assessments in that category (in parentheses) follow: standardized tests (43), curriculum-embedded tests (63), district objective-based tests (19), minimum competency tests (12), school-departmental, and/or grade-level tests (17), teacher-constructed tests (101), diagnostic instruments (11), and "other" evaluation techniques (75). The "other" category included such techniques as homework, worksheets, conferences, book reports, discussions, observations, etc.

As can be seen from the above frequencies, teacher-constructed tests and "other" evaluation techniques were cited most often by the teachers interviewed, a finding which is fairly consonant with Yeh's (1978) conclusion that curriculum-embedded tests and teacher-made tests are used to a much greater degree than standardized tests, but despite high frequency of testing, teachers are more likely to use personal observations and interactions with students than test results to assess student's progress. This latter point was not reflected in the freqiencies given above but it is possible that many of the teachers, and especially those at the elementary level, failed to mention many of the informal assessment activities that occur because they are used so frequently and are so much an integral part of the teaching process. This possibility influenced the manner in which we conceived and phrased items on the survey instrument so that the subject of informal assessment could be explored further.

The amount of time these assessment techniques take to prepare, administer, and/or grade was also explored. Again, as expected, a wide range of time spent on evaluation in the classroom was reported by the elementary and secondary teachers interviewed. However, on pursuing this matter it became apparent that teachers experienced difficulty in providing an exact estimate of time indices. This was due to a variety of reasons. For one, some teachers could simply not remember how long the tests took. More commonly, it was discovered that teachers allowed different students varying lengths of time to finish the tests and thus found it difficult to average the time amounts for all students. When asked about the informal techniques they used, teachers found it next to impossible to estimate the time they spent as many of the techniques were ongoing and/or overlapping.

Although the aforementioned difficulties were encountered during the interviewing process the teachers' reports gave some indication of the time devoted to evaluation. The teachers tended to be conservative in their estimates and when ranges of time were given for a particular assessment technique, we selected the midpoint of this time frame for analysis purposes.

The analysis of the data showed that the 22 elementary teachers interviewed spent an average of approximately 11 percent of their reading and math instructional/class time assessing their students. The 22 secondary teachers reported that about 24 percent of their English and math class time was spent on evaluation. The proportion of total classroom time given over to assessment was quite large for both elementary and secondary teachers; one to 64 percent for elementary and six to 75 percent for secondary.

At first glance it appeared on the average that the secondary teachers spent more time assessing their students than the elementary teachers. However, when looking at the responses concerning the type of assessments given, the vast majority of the secondary teachers' responses were for formal pencil-and-paper tests. Perhaps more formal testing is occurring at the secondary level than at the elementary grades because of the ages of the students involved and because the secondary teacher has less time for the use of informal techniques and/or observations. As the elementary teacher usually spends the full school day with the same group of students, he/she has more opportunity for informal evaluations and less need for the more formal ones. Also, because the informal techniques were not cited by the teachers as frequently as the more formal ones, the difference in the

percentages of time allotted to evaluation by the two sets of teachers was quite large.

The analysis also showed similar results for the total amount of time the teachers spent on evaluation. This total time includes the preparation, administration, and grading of tests/assessments. The elementary teachers reported on the average that 15 percent of their time (which includes instructional and non-instructional/preparation time) was spent on assessment while the secondary teachers spent 34 percent of their time on the same. The ranges reported by the elementary and secondary teachers were three to 56 percent and nine to 69 percent, respectively. Again, teachers' tendency not to report informal assessments and the use of many more formal evaluation techniques at the secondary level may account for some of the difference in the amount of time spent on assessment in elementary and secondary classrooms.

Range of Tests Administered

Fieldwork indicated that a wide range of test were being administered. For example, standardized tests, such as the Comprehensive Tests of Basic Skills (CTBS), the Metropolitan Achievement Test (MAT), Iowa Test of Basic Skills and of Educational Development (ITBS, ITED), etc., were administered in each school district visited.

Curriculum-embedded tests of various types were also given everywhere, but almost exclusively at the elementary grade levels. Most of the curriculum-embedded tests accompanied commercially-produced, elementary-grade series in math and reading. Among those given frequently were placement tests; the "unit" or "criterion" tests designed to assess achievement on a specific portion of the curriculum; and the "end of the book" tests (i.e., those the student took at the completion of a given reading or math "level").

Minimum competency tests were given in two of the districts. In one case they were district-developed and included four separate instruments assessing fundamental math skill and four assessing skills in the language arts. These tests were given at the high school level and passage of all eight was required for graduation. In the second district, an instrument developed by the state for administration to ninth grade students included the general domains of reading, mathematics, and writing. Its function was only diagnostic.

A statewide assessment measure was given annually in one district to a matrix sampling of students at certain elementary and high school levels. Individual student scores were not reported to schools, but aggregations by grade-level, school, and district were provided on various subskills in reading, mathematics, and writing.

District tests, district-constructed and mandated for use district wide, were part of the assessment picture in two of the three districts visited.

School-, departmental-, and/or grade-level tests were found in five school sites. One high school, for instance, had just developed and administered a writing sample in all grade levels. Departments in several high schools had teacher-developed mid-terms and finals for particular courses. And in two elementary schools in one of the districts, teams of teachers at particular grade levels constructed and gave common tests keyed to their social studies curriculum.

Diagnostic instruments were also employed largely, but by specialists such as remedial reading instructors, teachers of the "learning disabled" and "emotionally handicapped," and Title I program staff members. Almost

all of these were found in elementary schools.

Teacher-constructed tests, quizzes, and the like were, of course, extant in every site.

Other measures of student achievement were also prevalent in all classrooms. In the elementary grades, students' daily worksheets, classroom performance, along with homework and other assignments, were mentioned as ways of evaluating students' progress, These same types of "measures" were among those used by high school teachers. The latter also cited conferences with students, peer evaluation of classroom reports, oral quizzes and question-answer sessions, group discussions, and a wide variety of written assignments as assessment techniques.

Range of Reported Uses

Distinct patterns of use also grew out of fieldwork analysis, which suggested that test scores and other assessment results were used for a finite number of purposes across the sites visited. At the classroom level, there was little school-to-school or district-to-district variation in the range of uses respondents' reported. Eleven types of uses for assessment information were inductively derivable from the specific comments of educators interviewed. Recall that the uses listed below are those which individual respondents said they themselves made of test scores and other student assessment "data."

1) Referral to and/or placement in special programs, appropriate classes, appropriate "tracks," etc.

2) Within-classroom placement of students at appropriate levels in individualized programs, in reading or math groups, in occasional, temporary skills remediation groups, etc.

3) Planning instruction: "figuring out my class' strengths," "learning what the group needs," "getting feedback so I know what we have to go over again," "working with one of my grade-level groups of teachers to decide what areas they need to strengthen," etc.

4) Monitoring student's progress, "seeing how they're doing as we go along," "just getting a sense of whether they're learning anything."

5) Holding students accountable for doing assigned work, maintaining class discipline.

6) Assigning report card grades.

7) Certifying students' competency for promotion, high school graduation.

8) Counseling and advising students about how they are doing, about their preparation for future courses and academic goals, about their achievement, motivation potential, etc.

9) Informing parents of how their children are doing in regularly scheduled conferences, at "back-to-school" nights, special meetings, when problems arise.

10) Reporting to higher organizational levels within the district --to the principal, district office, the school board--on student achievement.

11) Comparing groups of students with others, judging how a class, school or district is performing relative to others

Patterns of Assessment Results Use

From the respondents' comments about how they used the results of particular tests and other assessments we developed a coding scheme to index the importance of particular results for particular purposes. This simple scheme depicted the use of a score or result for a given purpose as: (1) the sole information source used; (2) one of two or three major sources; (3) one of many sources; (4) a verification source, i.e., used ancillarily to check decisions or conclusions already reached based on other information sources; and (5) not used, simply administered.

Interview data from the 44 classroom teachers included 330 descriptions of how results of particular types of assessment were used.* They also included 21 statements that the respondents did not use results of types of measures that they administered.

As Table 1 indicates, teachers rarely used only one type of assessment information to make a given decision or accomplish a given purpose. Only 5.1 percent of the uses cited (including statements of non-use) were "sole source" uses, i.e. results used alone to make a given decision. In two-thirds of the cases, results from a particular type of assessment were used as one among many types of information employed for the particular purpose at hand.

Table 1

Overall Patterns of Assessment Results Use

	Functional Importance					
	Sole Source	One of Several Major Sources	One of many Sources	Verification Source	Not Used	Total
Instances Mentioned	18	65	237	10	21	351
	(5.1%)	(18.5%)	(67.5%)	(2.8%)	(6.0%)	(100%)

In short, it appeared that teachers were most likely to look at a variety of different kinds of information as they make the judgments, analyses, and reports they must make as part of their routine professional activities.

* Redundant uses for different tests of the same type were dropped out in collapsing the 346 tests/assessment means cited into the eight types of assessment listed earlier in this section.

Test information used as solo and major criteria: If most means of assessment provide information that is used jointly with others, which means do seem to provide information that functions as a sole or major criterion in teachers' activities? Table 2 provides an answer in overview.

Table 2

Types of Tests Used by Teachers
as Sole and Major Sources of Information for any Purposes

Test Type	Total* Citations All Levels	Count (Column %) Sole Source	Major Source	Total: Sole & Major (% total in Table)
Standardized	43	3 (33.3)	8 (7.7)	11 (13.2)
Curriculum Embedded	63	5 (27.8)	12 (18.5)	17 (20.5)
District Objective-Based	19	1 (5.6)	6 (9.2)	7 (8.5)
Minimum Competency+	12	0 (0.0)	0 (0.0)	0 (0.0)
Statewide Assessment	10	0 (0.0)	0 (0.0)	0 (0.0)
School/Department Grade-Level	17	0 (0.0)	9 (13.8)	9 (10.8)
Individual Teacher-Constructed	101	5 (27.5)	15 (23.1)	20 (24.1)
Diagnostic	11	0 (0.0)	0 (0.0)	0 (0.0)
Other	75	1 (5.6)	18 (27.7)	19 (22.9)
TOTALS	351	18	65	83 (100.0)

* Count of all instances in which test type was mentioned as used in any way, including "not used" category

\+ Minimum competency tests were used as the sole source for deciding whether students graduated from high school on one district, but this decision was not made by classroom teachers or other school-level practitioners.

From the above, a picture began to emerge of teachers drawing upon many types of assessment to do their routine instruction-related work. And the fieldwork data suggested that the types of assessment they use most frequently in this routine work tended to be those that are

- most immediately accessible to teachers and which provide most immediate results; those over which they have most control--can administer when they choose and can see the results promptly;
- Those which purport to serve functions isomorphic with the tasks teachers must routinely do; i.e., curriculum-embedded placement tests figure significantly in placement decisions; records of progress through a continuum for placement in a continuum; tests that teachers design or text publishers produce for measuring achievement on a unit of instruction for monitoring progress and grading students on that unit, etc.
- those which teachers deem to "cover" most exactly the content of the material they are teaching.

In short, those tests teachers see as linked most closely to the routine, practical activities of their everyday professional lives are those they use most often. Additionally, the phenomenological evidence of everyday experience with students plays an important role in teachers' assessments of them.

The single exception to this generalization appears to occur in the use of standardized tests. For the most part, teachers used these for general reference, to get an initial sense of how their new classes "look" relative to others, or as a normative reference point against which to guage progress--except, it seems, when they are required to do otherwise by district mandate.

Test information that is not used: In 21 instances, teachers said they did not use the results of one or another type of test that they gave. Ten teachers mentioned their non-use of standardized test results; seven

mentioned non-use of statewide assessment. In the case of the latter, teachers had no access to students' individual scores or results aggregated by class.

The above descriptions began to indicate some of the activities in which assessment results play a definitive or major role. Table 3 provides a comprehensive picture of the purposes for which they do so.

Table 3

Purposes for Which Teachers Use Various Types of Assessment Results as Sole and Major Information Sources

Count: Number of Citations

Purposes	Sole	Major	Total	(% Table Total)
Planning Instruction	1	9	10	(12.1%)
Referral/Placement: Special Program	4	5	9	(10.8%)
Within-Class Grouping & Individual Placement	7	18	25	(30.1%)
Holding Students Accountable for Work, Discipline	1	6	7	(8.9%)
Assigning Grades	0	9	9	(10.8%)
Monitoring Students' Progress	0	6	6	(7.2%)
Counseling & Guiding Students	5	8	13	(15.6%)
Informing Parents	0	1	1	(1.2%)
Reporting to District Officials, School Board, etc.	0	2	2	(2.4%)
Comparing Groups of Students, Schools, etc.	0	1	1	(1.2%)
*Certifying Minimum Competency	0	0	0	(0.0%)
TOTAL	18	65	83	

*Note: In one district visited, tests of minimum competency were required for high school graduation. Respondents, however, took this as obvious and rarely mentioned that they served in this way. When they did speak of the uses of minimum competency results, they described their uses for other purposes.

As Table 3 shows, test scores seemed to play an important role in student placement decisions. In 40.9 percent of the instances in which

teachers reported that they used assessment results as a sole criterion or a major criterion, the placement of learners was at issue. The use of scores as a major basis for in-class placement was especially frequent.

Summary. Most often, teachers seemed to consider the results of several types of assessment collectively in arriving at a particular decision or carrying out a particular activity. When they reported departing from this practice, it was more often in the direction of weighing test scores more heavily than in the direction of counting them less. (Citations of results as sole and major information sources equaled 23.6 percent of the total; citations of results not being used or used only in verification equaled 8.8 percent of the total.) The placement of students seemed to be an activity in which the results of one test or type of test may count more heavily than in others.

Relationships Between Types of Tests and Categories of Use

Table 4 summarizes the test type/use type relationships reported by both the elementary (n=22) and secondary (n=22) classroom teachers interviewed. The table indicates that the main uses of test and other assessment results include:

- Planning for instruction
- Grouping students and placing them at levels of individualized programs within classrooms
- Grading
- Monitoring students' progress, i.e., keeping track of how they are doing over time.

Table 4

Types of Tests and the Uses of Their Results

USES Counts: Elementary Secondary Cell Total	Type of Test: Standardized	Curriculum-Embedded	District Objectives-Based	Minimum Competency	Statewide Assessment	School, Department, Grade-Level	Teacher-Constructed	Diagnostic (Skills)	Other (Informal)	TOTAL
Planning Instruction	9 4 13	8 2 10	3 0 3	1 3 4	2 0 2	1 2 3	11 13 24	1 1 2	13 8 21	49 33 82
Referral/Placement:	9 2 11	0	0	0 1 1	0	0 2 2	2 1 3	0	2 4 6	13 10 23
Within Classroom Grouping & Individual Placement	4 0 4	18 0 18	5 0 5	1 2 3	0 1 1	1 3 4	2 4 6	6 0 6	11 3 14	48 13 61
Holding Students Accountable for Work, Discipline	0	3 0 3	0	0	0	0	4 4 8	0	2 0 2	9 4 13
Assigning Grades	0 1 1	14 3 17	1 0 1	0 1 1	0	0 5 5	15 17 32	1 0 1	7 1 8	38 28 66
Monitoring Students' Progress	0	14 0 14	4 0 4	0	0	0 2 2	10 8 18	1 0 12	10 2	39 12 51
Counseling & Guiding Students	1 2 3	0	2 0 2	0	0	0	2 8 10	1 0 1	4 2 6	10 12 22
Informing Parents	0	0	1 0 1	0	0	0	0	0	1 0 1	2 0 2
Reporting to District Officials, School Board, etc.	0	1 0 1	2 0 2	0	0	0	0	0	3 0 3	6 0 6
Comparing Groups of Students, Schools, etc.	1 0 1	0	1 0 1	0	0	0	0	0	1 0 1	3 0 3
Certifying Minimum Competence	0	0	0	0 1 1	0	0	0	0	0	0 1 1
TOTAL Use CITATIONS	24 9 33	58 5 63	10 0 19	2 8 10	2 1 3	2 14 16	22 55 101	10 1 11	51 20 74	217 113 330
Explicit Statements: "NOT USED"	5 5 10	0	0	1 1 2	0 7 7	1 0 1	0	0	0 1 1	7 14 21
Total Citations	29 14 43	58 5 63	19 0 19	3 9 12	2 8 10	3 14 17	46 55 101	10 1 11	54 21 75	224 127 351

Summary. The exploratory fieldwork indicated that the sample teachers most frequently drew on the results of three types of assessment. These are (1) their self-constructed tests, quizzes, and written assignments, (2) other assessment techniques that they devised or chose to seek out and use, such as class discussions, peer evaluations of work, conferences with students, talks with students' previous teachers, oral reading sessions, etc.; and (3) curriculum-embedded tests--those that come with district-made curriculum "packages" or commercially published texts, kits, and the like. They appeared to use each of these three types especially, but others as well, in accomplishing a variety of purposes. That is, teachers seemed to refer to each kind of assessment result for making a variety of judgments, just as they seemed to make a given decision by referring to a variety of assessment results. Principals seemed to engage in similar practice, although the test scores they used most often and the purposes for which they used them most frequently differed from those of the teachers. All this suggested, of course, that the national survey should examine patterns of test type/test use relationships. It should not assume simple one-to-one correspondences between a test score and a use.

Teachers most frequently cited test scores and other assessment results as serving them in four activities: Planning instruction, grouping and placing students in a continuum of objectives within the classroom, assigning grades, and monitoring students' progress over time. Counseling, guiding, and other use seemed to follow from the factors previously discussed.

A final point is worth noting again. Returning to Table 4, it is obvious that some activities for which teachers use student assessment results are relatively "under-mentioned." For instance, conferences with

parents are a routine part of teachers' work, especially at the elementary school level. A talk with any teacher about his/her students inevitably includes comparisons with students in other classes or school, students in previous years, and so forth. That these activities were cited relatively infrequently as uses of assessment was troublesome to us. In talking with teachers, however, it became evident that many of the practical tasks for which teachers use test information are, in fact, "transparent" to them. That is, they are so much a part of everyday life that they go un-noticed. They are treated, literally as unremarkable. That this is so is probably best illustrated by a comment made by a high school assistant principal in the first district visited, who explained in the same breath that they did not pay much attention to CTBS scores in his high school because the typical freshman entering the school was "two years, at least, below grade level."

This should serve as a caveat that Table 4, and the discussion which has followed from it, is not a complete picture of the frequency with which the teachers interviewed use test results for certain purposes. But, given the open-ended nature of the interviews, it is very likely a comprehensive picture, overall, of the kinds of uses that the test and other assessment results serve.

Pilot-testing of the National Survey Questionnaire

As further work in the design of our national survey, approximately 70 elementary teachers, secondary teachers, and principals in a Southern California school district responded early in 1981 to the draft versions of the elementary, secondary, and principal questionnaires. Of the 70 respondents, 36 were elementary teachers. At the time of preparing this paper, we were able to tabulate those elementary teachers' responses to see what similarities and disparities might exist between pilot-test work, the

fieldwork, and earlier CSE Study of Test Use.

Tables 5 and 6 summarize the pilot data regrading the number of types of tests used in the classroom and the number of administrations of those test types. Table 5 shows that teacher-constructed tests (line D) were the most common type of formal assessment for math and the second most common type of assessment for reading (behind commercial tests).

Table 6 indicates that teacher-made tests and quizzes are the most frequently administered type of classroom assessment. This corroborates Yeh's (1978) findings. However, a cautionary note must be sounded again regarding the reported number of administrations. While not exact, the estimates are approximate but still much higher than those given for other test categories.

One more point should be made about the pilot qiestionnaire results. The grand totals of both tables show more testing in reading than in math. This is at variance with other findings (see Yeh, 1978) and may be due to any of several factors. The final results of the Test Use Project will address this and other questions of interest regarding how tests are used.

Table 5

Types of Tests and Their Frequency of Use

	Reading/ Language Arts	Math
A. Tests Included with Commercially Published Curriculum Materials	67	24
B. District Developed Tests	39	15
C. Tests Developed by School/Department/Grade Level	13	18
D. Teacher Developed Tests And Quizzes	53	34
E. Written Assignments Used for Assessment	66	15
F. Miscellaneous Teacher Made Assessment	24	85
Grand Total	262	191

Table 6

Types of Tests and Their Number of Administrations Per Year

	Reading/ Language Arts	Math
A. Tests Included with Commercially Published Curriculum Materials	513	496
B. District Developed Tests	371	349
C. Tests Developed by School/Department/Grade Level	92	76
D. Teacher Developed Tests And Quizzes	1,330	1,302
E. Written Assignments Used for Assessment	1,214	278
Grand Total	3,520	2,501

THE DESIGN OF TESTING PROGRAMS WITH MULTIPLE AND COMPLIMENTARY USES

James Burry

Introduction

Some of the discussion on testing has recently begun to shift away from the purely social and psychological issues toward a concern with linkages between testing and instruction. This recent discussion views as one element in a broad set of assessment methods whose impact on and value for students and teachers is judged in terms of instructional practices. A prime question informing that judgment is -- Does a particular assessment method help in the day-to-day world of school and classroom decision making, especially in regard to diagnostic and prescriptive decisions about individuals and groups of students? A related question is --What assessment methods which are useful in classrooms and schools also have relevance for other levels of decision making in the educational system, decisions related to external accountability and to district, state, and federal policy concerns?

As instructional considerations have come into prominence, the dialogue over testing has become somewhat adversarial, with a great deal of the recent literature forming a series of position papers espousing the value of one kind of test over another, but offerring little empirical data (Lazar-Morrison, Polin, Moy, & Burry, 1980). A great deal of this debate is carried out by people outside the schools; the locus of the debate implicitly highlights the need to hear from teachers, principals, and other school people involved in daily classroom activities.

This paper makes a preliminary step toward explicating school peoples' points of view about the kinds of assessment that are useful for external

accountability concerns and for instructional decision making. More particularly, the paper will begin that explication by describing those elements in planning and design of assessment programs which seem to lead to the collection of information which has multiple and complementary uses. In providing this information, I will be describing the assessment practices in some of the schools in the three districts that were part of our exploratory fieldwork in CSE's Test Use Project -- a national survey of testing practices and test use in public elementary and secondary schools. The information I report here was collected in a series of interviews with teachers, counselors, and principals in the schools of these three districts. The sketch draws heavily on a content analysis of the responses of the people interviewed.

Content analysis of the taped transcriptions suggest that five factors seem to converge in the design of "exemplary" assessment programs:

(1) state testing policy and requirements

(2) coherence of school/district testing policy and requirements

(3) leadership in the instructional uses of assessment information

(4) locus of ownership in the assessment program

(5) recognition that no single test can serve (nor is intended to serve) the information needs of decision makers who reflect a variety of interests from broad program accountability to specific classroom practice.

While we had not intended fieldwork to provide a picture of "exemplary" test use, analysis of responses did suggest a tentative picture of how contextual factors may converge to make tests appear usable. As will be seen later, the district which seems to have the most successful program -- successful from the standpoint of reconciling or balancing external testing

requirements with school-level uses of testing -- assumes an organizational posture which has elements of centralism and diffusiveness. Put another way, this means that an organization and its constituent parts can be "loosely -coupled" in some regards and more tightly coupled in others. (For a discussion of these organizational causes and their effect in evaluation see Bank & Williams, 1981). This variable posture appears to lend itself to multiple uses of assessment information: uses which are central and concerned with external accountability and reporting requirements and uses which are spread out and reflect the decision needs of individual schools and classrooms. I am not suggesting that a balance of central authority and dispersed decision making is the only approach to the succesful design of an assessment program with multiple uses. But it appears to be the approach that has evolved, over time, in this particular district, and it seems to reflect not only organizational reality but the careful determination of various decision needs and specification of an assessment information system that will meet these needs.

Assessment programs often intend to provide information for use at local, state, and/or federal policy levels. Often the program will tend to emphasize the information needs of one of these levels to the exclusion of the others. Many assessment programs appear to be driven, or are perceived by the people in them, to be driven more by broad, external accountability than by concerns for classroom- and school-specific information. (This issue of external "linkages" is also dicussed in Bank & Williams, 1981.) Audiences associated with these external requirements often ask for assessment information that can be used to compare educational programs rather than to show the growth of individual pupils in terms of a specific set of

educational objectives. A school system which tends to respond more to the external audience than to others frequently relies on the collection and analysis of pupils' scores on a norm-referenced test. It may be criticized for lack of concern with individual students and their growth in a given classroom (no such system was discovered in the present study) might tend to rely more on criterion-referenced of objectives-based tests to provide information for diagnostic and prescriptive information. A school system taking this position might be subject to questions about the educational significance of the scores obtained on this kind of test -- What do they mean? Do they show whether the learning that has taken place is important or trivial? How do the scores obtained on these tests compare with the scores obtained on other kinds of tests?

A school sysyem might attempt to reconcile both kinds of information needs, to examine the operant assessment requirements, to investigate their own assessment needs, to determine which kinds of information will address the range of needs, to decide which kind of measure is most appropriate for generating the information addressing a particular decision area, to specify for its participants the intended uses of various measures, and thus design a coherent assessment program which is perceived to have a variety of overlapping uses.

One of the districts we spent time in appears to have developed this kind of assessment program. The two other districts we visited are trying to move in this direction, but still seem to be more concerned, or at least their teachers feel they are more concerned, with external accountability issues.

THE THREE SCHOOL DISTRICTS

District One

This school district, located in the urban northeast, has 24 elementary schools (kindergarten to grade 6 primarily; a few are K-8), 2 middle schools (grades 7-8), and 3 high schools (grades 9-12). Total enrollment is 27,000, with approximately 50% Black, 30% Hispanic, and 20% Anglo and other combined. The district has approximately 18 schools that are Title I eligible.

The state in which this district is located has a minimum competency testing program which is still in a formative stage of implementation. While no final determination had been made at the time data were collected, school district officials did not anticipate that the proficiency test would become a requirement for high school graduation. By the provisions of the state requirement, which focuses on "education, evaluation, and remedial assistance," all 9th graders are tested for proficiency. Any student scoring below a certain cut-score (established by the state) must receive remedial assisstance from the local school/district. The state required testing covers the areas of reading/language arts, mathematics, and also calls for a student writing sample.

Beyond the state required minimum competency testing program, the district has its <u>own</u> testing program, which is also in a formative stage of development. This district testing program deals with the areas of reading and communication arts, and includes the use of a locally developed criterion-referenced measure. This test is structured by grade, scope, and sequence, is intended to provide mastery data, and is administered by teachers and/or reading consultants. It becomes part of the student's permanent school record and follows him/her from grade to grade and school to

school. District officials anticipate that when this test has been fully developed, it will become part of the district's response to the state required minimum competency testing program.

As part of the district's required testing, the Metropolitan Achievement Test (MAT) is used in grades 2 through 8. It is administered every spring. At the high school level, the Comprehensive Test of Basic Skills (CTBS) is administered in the 11th grade.

The district test, which is accompanied by a specific curriculum, is supposed to be administered in all schools as part of an attempt to standardize the curriculum; this is apparently not happening in actual practice, however.

District Two

The second district we visited is located in an urban area in the southwest. This district has over 100 elementary schools, 20 junior high schools, and 14 high schools. Total district enrollment is a little over 100,000.

The state in which this district is located has a required minimum competency program for high school graduation. Local districts can use a state developed test or select/develop their own. This district has developed its own competency program to meet the state requirement. Among the tests in use in elementary schools are: CTBS; the state assessment program; the district competency test; and variable use of a range of curriculum-embedded tests and teacher observation and classroom interaction. Among the test in use in the high schools are: the state assessment program; district competency tests; CTBS; test associated with college entrance; and variable use of teacher constructed measures and classroom observation and interaction.

District Three

The third district visited, which demonstrated multiple and exemplary uses of assessment information, is located in a rural community in the mid-west. This district has seven elementary schools, three junior high schools, and one high school. Total district enrollment is a little over 5,000 students, of whom only 6 percent are minorities.

The state in which this district is located has no required minimal competency or proficiency testing. The only state requirement is that districts must identify students needs and set plans to meet desired levels of achievement.

Among the tests used are the Iowa Tests of Basic Skills (ITBS, grades 3-8), the Iowa Tests of Educational Development (ITED, grades 9-12), the Cognitive Abilities Tests (CAT, grades 1,3,6, and 9), district/school developed objectives-based tests, and curriculum-embedded tests.

Schools in this district also enjoy the resources of an Area Education Agency (AEA). One of the functions of this agency is to provide technical assistance to schools and individual teachers who have questions, problems, and needs in testing.

This district differs from the first and second on some important dimensions. In the third district, the fairly well accepted, district/school developed tests reduce the amount of time that teachers spend constructing and administering their own tests (especially at the elementary schools), thus freeing instructional staff for other tasks. There locally developed tests are largely seen as complementing the use of standardized tests, and serving different, though related decision needs. In addition, with greater acceptance of district testing there seems to be a clearer

sense among the teachers of both the "district" itself as an educational system and its testing policy and intentions, which teachers do not seem to see as threatening.

Much of the information provided by the respondents seem to reflect needs, issues, and concerns about three levels of decisions (Baker, 1978) that might need to be made on the basis of assessment information. Level 1, reflecting information needs to make decisions about individual students, is of prime concern among teachers, specialists, guidance counselors. Level 2, reflecting information needs to make decisions about groups of students within a school, is also of concern for some teachers, but somewhat more so among department chairpeople, grade level coordinators, and principals. Level 3, reflecting information needs to make decisions about groups across schools, is the concern of decision makers at LEA, SEA, federal levels, and the general public.

TEST USES/ISSUES IN DISTRICT ONE

In one of the schools in this district, an elementary school, respondents do not appear to value the district testing program. There is an impression that the administration, which had been recently appointed, was selected to stress the district program and the need for accountability at the level of the school. Respondents seem not to see the purpose or the relevance of the testing program. They do seem to be concerned with the kinds if tests available, their match with classroom curricular concerns, and the instructional unit at which the test has decision making relevance. Teachers here are largely concerned that the tests being used do not seem

to match their instructional concerns and related information needs. They see little coherence in the district/school testing policy.

In another elementary school in this district, the school administration and some of the curriculum and resource specialists seem to concern themselves to an extent with accountability (level 3) decisions, but the teachers do not seem overly concerned with this state of affairs. It appears that they not only go about the business of making their in-class and in-school (level 1 and 2) decisions, but also receive a level of expert assistance in making these decisions that was not encountered in the first school.

The third school visited in this district was a high school. Perhaps the most severe problem at the school is the fact that most of its students do not graduate. In an attempt to specifically pinpoint student deficiencies and make appropriate curriculum changes, the non-referenced test being administered -- the CTBS -- is a hope among staff that the district testing program (as well as improved use of department tests) will serve as student motivators and as a means to restructure the curriculu-.

District Summary

Several testing issues emerge in this district. First, the state-required testing program is still in a formative stage. The district testing program, which responds to state competency testing, is equally recent. The district program seems intended not only to serve the needs for competency testing but also to help standardize the curriculum district wide. At one school it is seen by teachers as no more than another accountability measure; if it has some instructional value, it is not seen by the

teachers. In this school, teachers seem to have little sense of district, or school, testing policy. Teachers seem to feel that required testing serves only level 3 decisions; it helps them not at all with level 1 and level 2 decisions and, indeed, may get in the way of teachers using measures of their own choice for these purposes.

In the second school, teachers seldom mentioned the district testing program. The teachers here perhaps understand the purposes of the program and so feel less threatened by it. On the other hand, they simply may not care either way if it does not get in the way of their classroom activities. One explanation is that concerns of the district testing porgram (and level 3 decisions) are seen in this school as the responsibility of the school administration and specialists. It appears that these specialists, some of whom are concerned about the amount of testing taking place, use the district measure not only for district concerns but also, where appropriate, to help classroom teachers with their internal level 1 and level 2 decisions.

In the third school, standardized tests administered in the past have served no purposes in instructional improvement. There is a distinct impression that the school is assuming a policy of "wait and see" in the hope that the new testing program will help them.

In general, the district testing program seems to suffer from lack of clear policy and guidelines; in only one of the elementary schools was there any sense of leadership in the instructional use of assessment information. It seems that at the high school a policy is emerging which may lead to a sense of ownership of the testing program.

TEST USES/ISSUES IN DISTRICT TWO

In one of the elementary schools in this district, a prime concern of the teachers is that tests will be used not only to monitor building progress, but also to evaluate teacher performance. The principal feels that if teachers beleive they will be evaluated on the basis of test scores, this is acceptable if that is what is required to achieve instructional improvement.

In the second school visited, a high school, the impact of minimal competency testing and the time devoted to this testing has had a profound influence both on teacher attitude toward testing and also toward the uses they make of other kinds of tests.

In the third school visited, also a high school, the impact of minimal competency testing was felt to be equally high, influencing not only the amount of testing taking place but also the content of instruction in the classroom.

District Summary

The advent of minimum competency testing has had an observable and, from the standpoint of some respondents, a negative effect on regular classroom instruction and the kinds of resource options made available to teachers. While the effect seems to be more pronounced at the high schools, it also seems to have a bearing on the policies of elementary schools visited.

In may respects, teacher concern for amount of testing, kinds of tests administered, and the uses to which they are put echo the kinds of responses encountered in the first district visited. This is especially true with respect to minimal competency testing.

TEST USES/ISSUES IN DISTRICT THREE

In one of this district's elementary schools, while there were some teacher-perceived problems with testing, teachers seemed to view tests as a more useful decision-making tool than was the case in the first two districts. The test selection/development/use inservice offered in this district appears to strongly influence teacher acceptance and use of test results. Of equal importance, however, are the services offered by the AEA, a kind of teachers center in which advice, technical assistance, and actual tests can be constructed/selected by teachers.

Another factor that appears to influence teacher use of tests is the atmosphere in which testing policy is conveyed. The district and school administration seem to set broad test information requirements intended to serve both external accountability and internal instructional improvement needs, in which departments and teachers have several options.

One of the respondents in the first school visited described the history of the district's approach to testing and the role of centralized training and technical assistance. As a media specialist responsible for providing "teachers with the materials they need to teach kids," several years ago he developed an interest in computer assisted instruction. His interest in CAI led to using local computer services for test scoring and data analysis. This led to a district interest in "computer analysis rather than hand scoring, to give you a better idea (of) where the kids are ... You don't have the time or expertise in the classroom, generally, to do that; the computer does it in one fell swoop." This quick and accurate scoring service, covering all the various kinds of tests used, is now

available to any teacher in the district. Over the years, further, the link from CAI to test scoring and analysis has led to a further computer application. That is, teachers have gradually developed large banks of educational objectives, have written or adapted hundreds of tests items written at varying levels of difficulty, and can now resort to the computer files to call out a particular kind of test for a particular instructional purpose. Over the years it appears that local teacher involvement, with technical assistance and leadership from the AEA and district officials, has led to a greater degree of test sophistication and test use among teachers than was the case in district one and two schools.

Therefore, while some teachers expressed concerns about such problems as the lateness of receiving results of the standardized test as well as its relevance for some classroom objectives, these criticisms did not carry over to testing in general. Indeed, some of the tests used are seen as invaluable for both teachers and students. Tests also seem to be used as instructional motivators whose results are discussed by teachers and students as one more source of diagnostic information. The link between testing policy and test use seems clearer than in the first two districts. In the third district teachers seem to feel the testing program is in part their own, to be used for their level 1 and 2 classroom decisions as well as for school and district accountability matters, and to be tempered by teachers' professional interactions with their students.

The second school visited, also an elementary school, appeared similar to the first in terms of uses of assessment information. The norm-referenced test in use -- the ITBS -- does not appear to receive a great deal of emphasis for classroom decisions, although it is useful to the

administration in making decisions about buidling-level effectiveness.

District developed and validated tests do appear to be weighed heavily for certain kinds of within-class decisions as well as for teacher self-monitoring. For many of these decisions, further, teachers also rely on less formal means of assessment in the interests of making the best instructional decisions.

The third school visited was a high school. Here some of the school staff interviewed seem knowledgeable (in some cases, almost expert) in matters of testing and test use, in the math department. Indeed, the school administration hopes that a model of the math department will eventually transfer to other departments. To be effective, however, they feel this must occur naturally with no direct interference from the administration.

In this school, the principal and associate principal emphasize the crucial role of the district in sponsoring within-school and centralized opportunities for technical assistance in testing. This school also seems to exemplify the best uses of certain knids of tests. In terms of the ITED, its use, as seen by the school administration, is as follows: "We need at least one outside measure, something outside of our own control .. so we can just have a benchmark ... that we can compare with" in terms of school-level performance. Beyond that, item analysis of ITED scores might lead to discussion between the associate principal and a department chair if test score trends are poor in certain areas. "Should this indication lead to course modification? Adding something to instruction? Do instructors want to add this area to instruction? Do they want to leave it out because they don't think it's important?" This kind of discussion suggests a measure of department autonomy or, at least, negotiated decision-making.

In this school in general, and in the math department in particular,

the school-developed measures appear to be accepted and used by teachers. Departmental autonomy in testing and the inservice and technical assistance made available appear to have stimulated local development of tests that are quickly accessible, fit teachers' practical needs, and have high content and classroom relevance. Standardized tests are primarily used by the school administration, and seem to be viewed neither as a threat nor as an unnecessary burden by the teachers.

District Summary

This district clearly has a different approach to testing and testing policy than the first two. It appears that the district establishes broad policy for schools, and the schools in turn, set broad policy for the instructional teams in the elementary schools and the departments in the high schools. Test administration, quality, and level 1 and 2 uses are also focused at the level of team or department. In addition, both the district central office and staff of the AEA provide active leadership in the development of tests and their instructional uses. Policy is clear, though flexible; it seems to reflect an organizational system whose units can "couple" or "decouple" as described in Bank and Williams (1981). A great deal of the testing appears to be "owned" by the school unit of concern--team or department. While teachers seem less likely to rely greatly on the ITBS and the ITED, counselors are available to help interpret these scores and place them in the larger assessment context for individual teachers.

Teacher knowledge of tests and testing appears to be greater than in the first two districts. There also appears to be more inservice and there is certainly much more technical assistance available in the third

district. This seems to have led to the development of tests of higher quality which apparently have marked instructional relevance for the teachers. The testing situation appears to come close to the ideal. That is, the overall testing program:

- offers tests oriented to classroom teachers
- permits teachers to use tests so as to meet their practical activities and exigencies
- does not force teachers to emphasize tests that do not fit their practical demands
- permits teachers to administer/use a variety of tests
- is sensitive to the practical matters of teaching

In this district, further, the merits of different kinds of measures are not discussed in an adversarial setting. Instead, the teachers, principals, and district officials seem to accept the need for and value in generating information that will paint the big (norm-referenced) picture, that will provide a wide angle view about groups and programs. They don't over-emphasize this picture. They also accept the need to generate information about the individual students and classrooms (criterion-referenced or objectives-based) that together make up the big picture. They don't over-emphasize the value of this picture either.

They seem to be using the right kind of test to get the larger aggregate picture, and a series of other equally appropriate measures, to get a variety of snapshots with a closer focus and with greater detail, of the separate parts of the picture. The district, the central figure, has supplied the camera -- the means to get different pictures -- and takes the kind of shot with the degree of resolution it needs. The schools and classrooms use the same camera, but they select a kind of film that meets their needs, and then choose an angle, focus, and degree of resolution

sensitive enough to get the series of shots that they need. The end result seems to be a montage reflecting different degrees of instructional progress among different aggregates of students at varying points in time. The whole is pleasing esthetically and technically.

REFERENCES

Baker, E.L. Achievement testing in urban schools: New numbers. Paper presented ar CEMREL Conference on Urban Education, 1978. (Also CEMREL Monograph on Urban Education, St. Louis, Missouri, CEMREL, 1980).

Bank, A. & Williams, R. Evaluation in school districts: Organizational perspectives. CSE Monograph Number 10. Los Angeles: Center for the Study of Evaluation, 1981, in press.

Lazar-Morrison, C., Polin, L. Moy, R., & Burry, J. A review of the literature on test use. CSE Report No. 144, Center for the Study of Evaluation, University of California, Los Angeles, 1980.